

Quality Assessment in Complex Classification Workflows

Qiong Bu & Elena Simperl

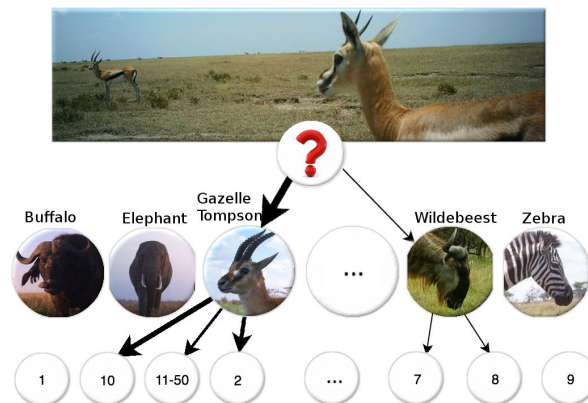
qb1g13@soton.ac.uk & E.simperl@soton.ac.uk

School of Electronics and Computer Science

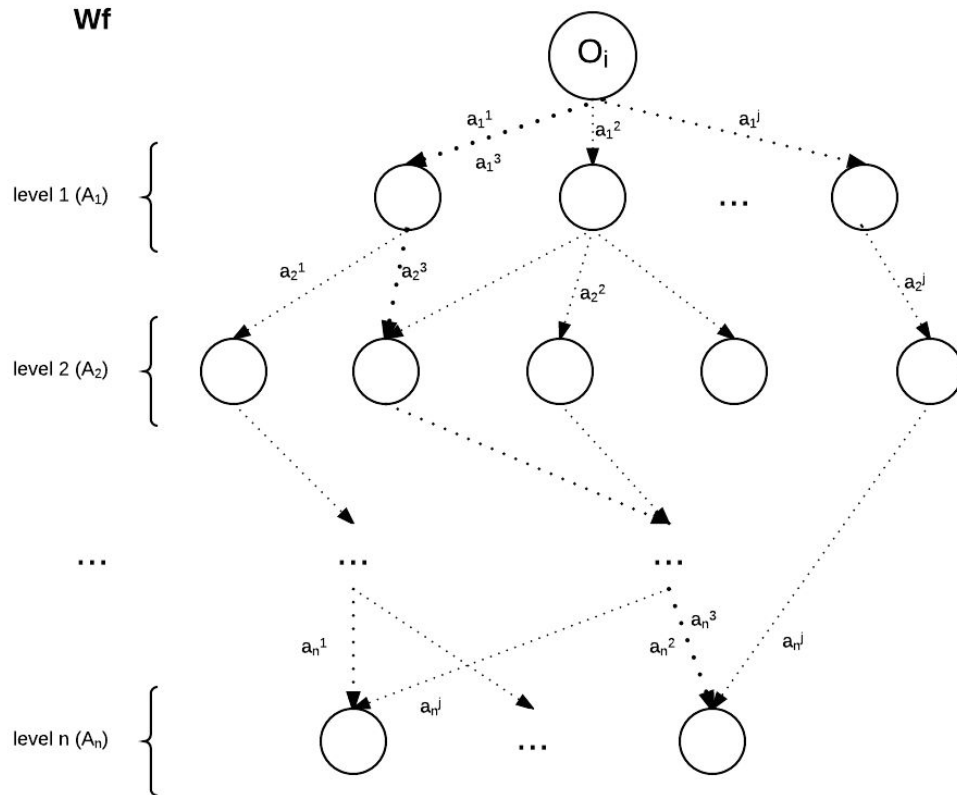
Goal

Understanding the Relationship between Task Design and Crowd Performance

- Explore task features that affect the quality of crowd work
- Define models that capture complex classification workflows
- Define aggregation method to infer answers in complex workflows



Approach - Conceptualization



a_x^y : y user's annotation for level x ($a_x^y \in A_x$, in which A_x is the set of available options at level x)

A classification task as a path search problem in a workflow modelled as DAG

The Correct Labelling Problem:

Input:

- object O_i
- a workflow graph Wf
- a set of labels for object O_i -- $ASET(O_i)$
- (optional) all labels $ASET(O \times U)$

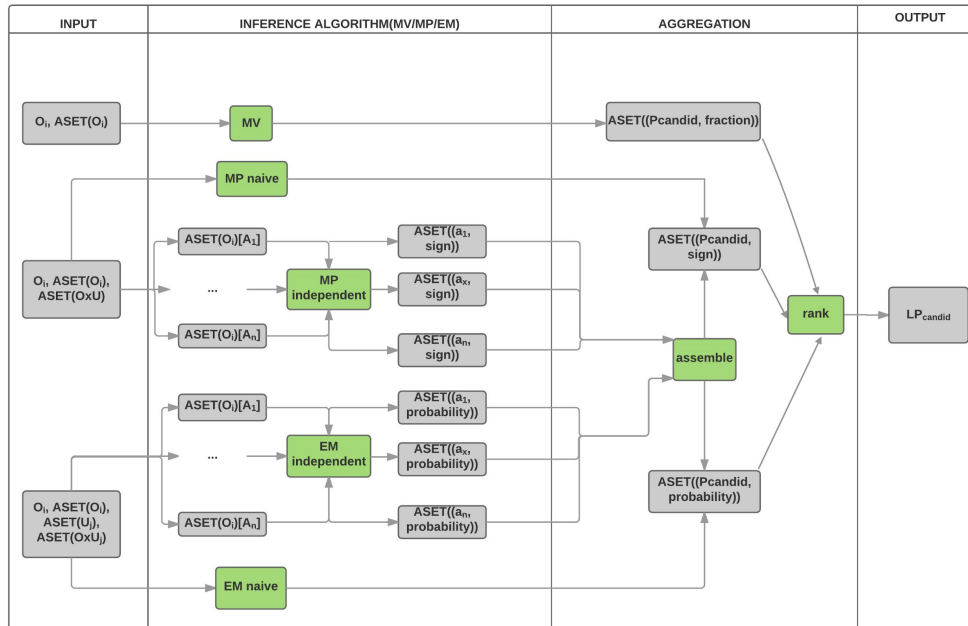
Output:

- ranked list of paths LP_{candid} as candidate true labels for O_i

Approach -Workflow-aware Inference

Algorithm 1 Majority voting

- 1: **procedure** FINDUNIQUELABEL($ASET(O_t)$)
- 2: $S_{unique} \leftarrow \{a_{p1} \dots a_{pz}\}$, where $S_{unique} \subseteq ASET(A)$ and $\{a_{pz} \in ASET(A) \text{ and } z \leq |ASET(O_t)|\}$;
- 3: **for** $a \in S_{unique}$ **do**
- 4: $Frac_p \leftarrow count(a) \div |ASET(O_t)|$
- 5: **return** $L_{candid} = \{(a_{p1}, Frac_{p1}) \dots (a_{pz}, Frac_{pz})\}$;



Algorithm 2 Message passing

- 1: **procedure** INITIALISATION($y_{j \rightarrow i}$)
- 2: **for** $(i, j) \in ASET(O \times U)$ **do**
- 3: Initialise $y_{j \rightarrow i} \sim \mathcal{N}(-1, 1)$;
- 4: **procedure** ITERATION(k_{max})
- 5: **for** $k \in \{1, \dots, k_{max}\}$ **do**
- 6: **for** $(i, j) \in ASET(O \times U)$ **do**
- 7: $x_{i \rightarrow j}^k \leftarrow \sum_1^{j-} A_{ij} \times y_{j \rightarrow i}^{k-1}$ ($j^- \neq j$);
- 8: **for** $(i, j) \in ASET(O \times U)$ **do**
- 9: $y_{j \rightarrow i}^k \leftarrow \sum_1^{i-} A_{ij} \times x_{i \rightarrow j}^k$ ($i^- \neq i$);
- 10: **procedure** ESTIMATE BASED ON WEIGHTED SUM(O_t)
- 11: **for** $i \in ASET(O)$ **do**
- 12: $x_i \leftarrow \sum_1^j A_{ij} \times y_{j \rightarrow i}^{k_{max}-1}$;
- 13: **return** $L_a = [sign(x_i)]$

Algorithm 3 Expectation maximisation

- 1: **procedure** INITIALISE(p_a)
- 2: probability of 'a' being the true label for object $O_t(a \in ASET(A))$:
- 3: $p_a \leftarrow count(a) \div |ASET(U_j)|$
- 4: **while** not converged **do**
- 5: Estimate error rate for labeller j:
- 6: $\theta_{aa}^j \leftarrow \lambda_{aa}^j + \sum_{o \in ASET(O \times U_j)} p_a \times I(a_o^j = a^-)$
- 7: Estimate confusion matrix:
- 8: $e_{aa}^j \leftarrow \theta_{aa}^j \div \sum_q \theta_{aq}^j$ (q is the accuracy of user j)
- 9: Estimate class priors:
- 10: $pr_a \leftarrow \sum_o p_a^o \div |ASET(O)|$
- 11: Calculate class probability for object O_t :
- 12: $p_a \leftarrow pr_a \prod_{j \in ASET(U_j)} \prod_m (e_{am}^j) I(a^j = m) \div \sum_q pr_q \prod_m (e_{qm}^j) I(a^j = m)$

Evaluation

Data

- Representative sample of classifications: classification tasks from Citizen Science projects

ZOONIVERSE
REAL SCIENCE ONLINE

STARS4ALL

- Three datasets:



Metrics

- Mean Average R-Precision & Averaged Discounted Cumulative Gain & Accuracy
- Comparison between algorithms (naive vs. workflow-aware)
- Aggregation result is compared to gold standard

Preliminary results

- Workflow-aware aggregation performs better in complex workflows

MV	MP	EM	Accuracy (SG,GZ,DS)
0.776, 0.871, 0.910	0.017, 0.002, 0.560	0.645, 0.516, 0.575	Naive
	0.756, 0.572, 0.485	0.883, 0.857, 0.900	Workflow-aware

- No single algorithm outperforms the others on all data
- Quality assessment in complex workflows is not agnostic of workflow properties
- Some characteristics of corresponding objects make it impossible to reliably infer correct label which should be further considered apart from the workflow structure

Acknowledgments

Thank you for the following funders of my student travel grant!

ACM SIGIR

Institute for Advanced Studies (IAS), University of Surrey, UK

Surrey Centre for Cyber Security (SCCS), University of Surrey, UK

School of Computer Science and Informatics, Cardiff University, UK

EU FP7 Marie Curie Initial Training Network ESSENCE

EU H2020 project QROWD

EU H2020 project Stars4All

EU H2020 project LETS CROWD

Singapore-UK project COMMANDO-HUMANS

Singapore-UK project STRICT

THANK YOU

Qiong Bu & Elena Simperl

qb1g13@soton.ac.uk & E.simperl@soton.ac.uk